

NEW THREE-TERM CONJUGATE GRADIENT METHOD FOR SOLVING UNCONSTRAINED OPTIMIZATION PROBLEMS

KHALIL K. ABBO¹ & LINDA A. ABDUL WAHID²

¹Department of Mathematics, College of Basic Education, University of Telafar, Iraq

²Department of Mathematics, College of Comp Science and Mathematics, University of Mosul, Iraq

ABSTRACT

In this paper we used, the Dai-Yuan nonlinear conjugate gradient (DYCG) method, to the three-term conjugate gradient (TDYCG) method and the derivation of the method, based on the Perry's conjugacy condition. The global convergence was proved, with Wolfe line search. Numerical experiments are reported and shows the presented method outperforming, some other three-term CG methods.

KEYWORDS: Three-Term Conjugate Gradient, conjugate gradient algorithms, Of quasi-Newton methods

Received: Jun 20, 2017; **Accepted:** Jul 10, 2017; **Published:** Jul 18, 2017; **Paper Id.:** IJMCARAUG20173

INTRODUCTION

We consider the unconstrained optimization problem:

$$\min f(x), \quad x \in R^n \quad (1.1)$$

Where $f: R^n \rightarrow R$ is a continuously differentiable function and bounded below the nonlinear conjugate gradient method is an iterative method whereby at the $(k+1)^{th}$ iteration x_{k+1} is given by:

$$x_{k+1} = x_k + \alpha_k d_k \quad (1.2)$$

Where d_k denotes the search direction and α_k is the step length. The search direction d_{k+1} is calculated by

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k, \quad d_1 = -g_1, \quad k=0,1,\dots \quad (1.3)$$

In (1.3) the quantity $g_k = \nabla f(x_k)$ denotes the gradient of f at x_k , while β_{k+1} is scalar which can be defined by

$$\begin{aligned} \beta_{k+1}^{FR} &= \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, \quad \beta_{k+1}^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}, \quad \beta_{k+1}^{CD} = -\frac{g_{k+1}^T g_{k+1}}{d_k^T g_k} \\ \beta_{k+1}^{PR} &= \frac{g_{k+1}^T y_k}{g_k^T g_k}, \quad \beta_{k+1}^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \end{aligned} \quad (1.4)$$

Or by other formula (e.g. see [1,2,7]).The corresponding method is respectively called FR(Fletcher-Reeves[6]),DY(Dai-Yuan[3]),CD(conjugate-Decent[7]),PR(Polak-Ribiere[11]),HS(HestenesStiefel[9]),respectively.

Where $\|\cdot\|$ means the Euclidean norm. The line search in conjugate gradient algorithms is often based on the general Wolfe conditions [13,14]

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k d_k \quad (1.5)$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \quad (1.6)$$

Where d_k is a descent direction and $0 < \delta < \sigma < 1$. However, for some conjugate gradient algorithms, stronger Wolfe (SW) condition used, defined by equation (1.5) and:

$$|g_{k+1}^T d_k| \geq -\sigma g_k^T d_k \quad (1.7)$$

Are needed to ensure the convergence and to enhance the stability.

The pure conjugacy condition is represented by the form

$$d_{k+1}^T y_k = 0 \quad (1.8)$$

For nonlinear conjugate gradient methods. The extension of the conjugacy condition was studied by Perry [10], he tried to accelerate the conjugate gradient method by incorporating the second-order information into it specifically, he used the secant condition

$$H_{k+1} y_k = s_k \quad (1.9)$$

Of quasi-Newton methods, where asymmetric matrix H_{k+1} is an approximation to the inverse Hessian. For quasi-Newton methods, the search direction d_{k+1} can be calculated in the form

$$d_{k+1} = -H_{k+1} g_{k+1} \quad (1.10)$$

By (1.9) and (1.10) the relation

$$d_{k+1}^T y_k = -(H_{k+1} g_{k+1})^T y_k = -g_{k+1}^T (H_{k+1} y_k) = -g_{k+1}^T s_k \quad (1.11)$$

Holds By taking this relation into account, Perry replaced the conjugacy condition (1.8) by the condition

$$d_{k+1}^T y_k = -g_{k+1}^T s_k \quad (1.12)$$

Recently Zhang, Zou and Li [15, 16], proposed the following three-term conjugate gradient methods

$$d_{k+1} = -g_{k+1} + \beta^{FR} d_k - \gamma^{(1)} g_k \quad (1.13)$$

$$d_{k+1} = -g_{k+1} + \beta^{PR} d_k - \gamma^{(2)} g_k \quad (1.14)$$

$$d_{k+1} = -g_{k+1} + \beta^{HS} d_k - \gamma^{(3)} g_k \quad (1.15)$$

$$d_{k+1} = -g_{k+1} + \beta^{DY} d_k + \gamma^{(4)} g_k \quad (1.16)$$

Where

$$\gamma_{k+1}^{(1)} = \frac{g_{k+1}^T d_k}{\|g_k\|^2} = \gamma_{k+1}^{(2)} = \frac{g_{k+1}^T d_k}{g_k^T g_k}, \gamma_{k+1}^{(3)} = \frac{g_{k+1}^T d_k}{y_k^T d_k}, \gamma_{k+1}^{(4)} = \frac{-g_{k+1}^T g_k}{y_k^T y_k} \quad (1.17)$$

The rest of the paper is organized as follows: In section 2, we presented a new three-term conjugate gradient (TKLCG) method, which was obtained by modification of the Khalil-Linda CG method. Section 3, presents the global convergence of the proposed algorithm. In section 4, some numerical results and comparisons with some other three-term conjugate gradient methods, were presented.

NEW THREE-TERM CG-METHOD (TKLCG)

In this section, we derive a new three-term conjugate gradient algorithm, for unconstrained optimization which is generalization to, the Dai-Yuan CG method. Consider the following three-term search direction

$$d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k^{DY} s_k + \gamma_{k+1}^{KL1} y_k \quad (2.1)$$

If we use inexact line search and Perry conjugacy condition (1.12), we get from (2.1) the following value for θ_{k+1}

$$\theta_{k+1} = 1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} \quad (2.2)$$

Therefore our new three-term (TKLCG) conjugate gradient algorithm can be defined as

$$d_{k+1} = -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k}\right) g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} s_k - \frac{g_{k+1}^T g_k}{y_k^T y_k} y_k \quad (2.3)$$

We see from (2.3), for quadratic convex function with exact line search the above search direction reduces to the Fletcher CD method i.e.

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{-g_k^T s_k} s_k \quad (2.4)$$

(TKLCG) Algorithm

Step1. Given $x_1 \in R^n, \varepsilon > 0, d_1 = -g_1; k = 1$

Step2. If $\|g_{k+1}\| \leq \varepsilon$, stop, else go to Step 3

Step3. Find α_k satisfying Wolfe condition (1.5) and (1.6).

Step4. Compute a new iterative x_{k+1} by $x_{k+1} = x_k + \alpha_k d_k$.

Step5. Compute β^{DY} , γ_{k+1} and θ_{k+1} from (2.2) $d_{k+1} = -\theta_{k+1} g_{k+1} + \beta^{DY} s_k + \gamma_{k+1}^{KL1} y_k$ and set $k=k+1$ go to step 2.

3. DESCENT PROPERTY AND GLOBAL CONVERGENCE ANALYSIS

Next we will show that our three-term CG (2.3)-method satisfies the descent property and global converges.

Theorem 3.1. Let $\{x_k\}$ and $\{d_k\}$ be generated by the equation (1.2),(2.3) and α_k satisfies Wolfe line search conditions (1.5) and (1.6), then $d_k^T g_k < 0$ hold for all $k \geq 1$.

Proof. The conclusion can be proved by induction. When $k = 1$, we have $d_1^T g_1 \geq -\|g_1\|^2 < 0$

Suppose that $d_k^T g_k < 0$ hold for all k . From (2.3) we have

$$d_{k+1}^T g_{k+1} = -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k}\right) g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} g_{k+1}^T s_k - \frac{g_{k+1}^T g_k}{y_k^T y_k} g_{k+1}^T y_k$$

$$d_{k+1}^T g_{k+1} = -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k}\right) \|g_{k+1}\|^2 + \frac{g_{k+1}^T s_k}{y_k^T s_k} \|g_{k+1}\|^2 - \frac{g_{k+1}^T g_k}{y_k^T y_k} g_{k+1}^T y_k$$

$$d_{k+1}^T g_{k+1} = -\left(\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k}\right) - \frac{g_{k+1}^T s_k}{y_k^T s_k}\right) \|g_{k+1}\|^2 - \frac{g_{k+1}^T g_k}{y_k^T y_k} g_{k+1}^T y_k$$

$$d_{k+1}^T g_{k+1} = -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} - \frac{g_{k+1}^T s_k}{y_k^T s_k}\right) \|g_{k+1}\|^2 - \frac{(g_{k+1}^T g_{k+1} - g_{k+1}^T g_k)}{y_k^T y_k} g_{k+1}^T g_k$$

$$d_{k+1}^T g_{k+1} = -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} - \frac{g_{k+1}^T s_k}{y_k^T s_k}\right) \|g_{k+1}\|^2 - \frac{g_{k+1}^T g_k}{y_k^T y_k} \|g_{k+1}\|^2 + \frac{(g_{k+1}^T g_k)^2}{y_k^T y_k}$$

$$\because g_{k+1}^T g_k \leq \|g_{k+1}\|^2$$

$$\because d_{k+1}^T g_{k+1} \leq -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} - \frac{g_{k+1}^T s_k}{y_k^T s_k}\right) \|g_{k+1}\|^2 - \frac{g_{k+1}^T g_k}{y_k^T y_k} \|g_{k+1}\|^2 + \frac{g_{k+1}^T g_k}{y_k^T y_k} \|g_{k+1}\|^2$$

$$d_{k+1}^T g_{k+1} \leq -\left(1 + \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} - \frac{g_{k+1}^T s_k}{y_k^T s_k}\right) \|g_{k+1}\|^2$$

By Wolfe condition $y_k^T s_k > 0$ and let $0 \leq \frac{g_{k+1}^T s_k}{g_{k+1}^T y_k} - \frac{g_{k+1}^T s_k}{y_k^T s_k} \leq 1$, then we can deduced that

$d_{k+1}^T g_{k+1} < 0$ holds for all $k \geq 1$.

To prove the global convergence we need the following assumption [8]

H1. The objective function f is bound below in the level set $\Omega = \{x \in R^n : f(x_1) \leq f(x)\}$ where

x_1 the starting point is.

H2. f Is continuously differentiable in a neighborhood N of Ω and its gradient g is Lipchitz continuous, there exist $L > 0$ such that

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \forall x, y \in N \quad (3.1)$$

H3. f is uniformly convex function, then there exists a constant $\mu > 0$ such that

$$y_k^T s_k \geq \mu \|s_k\|^2 \quad \text{and} \quad \mu \|s_k\|^2 \leq y_k^T s_k \leq L \|s_k\|^2 \quad (3.2)$$

On the other hand, under assumption [10], it is clear that there exist positive constant B such

$$\|x\| \leq B, \forall x \in \Omega \quad (3.3)$$

$$\underline{\gamma} \leq \|g(x)\| \leq \overline{\gamma}, \forall x \in \Omega \quad (3.4)$$

Lemma (1) [8]

Suppose that Assumption [10] and equation (3.3) hold .consider any conjugate gradient method in from (1.2) and

(1.3), where d_k is a descent direction and α_k is obtained by the strong Wolfe (SW) .If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \quad (3.5)$$

Then we have

$$\lim_{k \rightarrow \infty} \left(\inf \|g_k\| \right) = 0 .$$

More details can be found in [1,4and 12]

Theorem (1)

Suppose that, assumption [8], equation (3.3) and the descent condition hold. Consider a conjugate gradient method in the equation (2.3) where α_K is computed from Wolfe line search condition (1.5) and (1.6) If the objective function is uniformly on set Ω , then $\lim_{k \rightarrow \infty} \left(\inf \|g_k\| \right) = 0$.

Proof

$$d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k^{DY} s_k + \gamma_{k+1}^{KL1} y_k \quad (3.6)$$

Where

$$\|d_{k+1}\|^2 = \left\| -\theta_{k+1} g_{k+1} + \beta_k^{DY} s_k + \gamma_{k+1}^{KL1} y_k \right\|^2 \quad (3.7)$$

Using Cauchy Schwartz together with inequality (3.3) we get

$$\|d_{k+1}\|^2 \leq \theta_{k+1} \|g_{k+1}\|^2 + \beta_k^{DY} \|s_k\|^2 + \gamma_{k+1}^{KL1} \|y_k\|^2 \quad (3.8)$$

$$\|d_{k+1}\|^2 \leq \theta_{k+1} \|g_{k+1}\|^2 + \beta_k^{DY} \|s_k\|^2 + \gamma_{k+1}^{KL1} L \|s_k\|^2 \quad (3.9)$$

Let $M = \|s_k\|^2$ then we get

$$\|d_{k+1}\|^2 \leq \theta_{k+1} \|g_{k+1}\|^2 + \beta_k^{DY} M + \gamma_{k+1}^{KL1} L M \quad (3.10)$$

$$\|d_{k+1}\|^2 \leq \theta_{k+1} \|g_{k+1}\|^2 + M \left(\beta_k^{DY} + \gamma_{k+1}^{KL1} L \right) \quad (3.11)$$

By Using Assumption [8] then we get

$$\|d_{k+1}\|^2 \leq \theta_{k+1} \bar{\gamma}^{-2} + M \left(\beta_k^{DY} + \gamma_{k+1}^{KL1} L \right) \quad (3.12)$$

Let $F = \theta_{k+1} \left(\gamma^{-2} \right)^2 + M \gamma^{-2} \left(\beta_k^{DY} + \gamma_{k+1}^{KL1} L \right)$ then we get

$$\|d_{k+1}\|^2 \leq F \frac{1}{\gamma} \quad (3.13)$$

$$\sum_{k=1} \frac{1}{\|d_{k+1}\|^2} \geq \frac{1}{F} \gamma^{-2} \sum_{k \geq 1} 1 = \infty \quad (3.14)$$

By Using Lemma (1) then we get

$$\lim_{k \rightarrow \infty} \left(\inf \|g_k\| \right) = 0 \quad (3.15)$$

4. NUMERICAL RESULTS AND COMPARISONS

In this section, we compared the performance of new formal θ_{k+1} , and developed a new three-term method of conjugate gradient method, to other classical conjugate gradient method Dai-Yuan (DY) and Khalil-linda (KL1) algorithms). We have selected 20 large scale unconstrained optimization problems, for each test problems taken from Andrie, (2008) [2]. For each test function, we have considered numerical experiments with the number of variables $n=100, \dots, 1000$. These two new versions were compared with well-known conjugate gradient algorithm, Dai-Yuan (DY) and Khalil-linda (KL1) algorithms. All these algorithms were implemented with standard Wolfe conditions (1.5) and (1.6). In all these cases, the stopping criteria is the $\|g_k\| = 10^{-6}$. All codes were written in double precision, FORTRAN language, with F77 default compiler settings. The test functions usually start the point standards initially summarizing numerical results for 730 functions, recorded in the Table (4.1). The results for the 730 functions, explained in Figure (1), (2) and (3), the performance profile, by Dolan and More' [5] was used, to display the performance of the Generalized Dai-Yuan Nonlinear Conjugate Gradient algorithm, with Dai-Yuan (DY) and Khalil-linda (KL1) algorithms. Define $P=730$ as the whole set of N_p test problems and $S=3$ the set of the interested solvers. Let $l_{p,s}$, be the number of objective function evaluations, required by solver S for problem P. Define the performance ratio as

$$r_{p,s} = \frac{l_{p,s}}{l_p^*} \quad (4.1)$$

Where $l_p^* = \min \{ l_{p,s} : s \in S \}$. It is obvious that $r_{p,s} \geq 1$ for all p,s. If a solver fails to solve a Problem, the ratio $r_{p,s}$ is assigned to be a large number M. The performance profile for each solver S is defined as the following cumulative distribution function for performance ratio $r_{p,s}$,

$$\rho_s(\tau) = \frac{\text{size}\{p \in P : r_{p,s} \leq \tau\}}{n_p} \quad (4.2)$$

Obviously, $p_s(1)$ represents the percentage of problems, for which solver S is the best. See figure [5] for more details, about the performance profile. The performance profile can also be used to analyze, the number of iterations, the number of gradient evaluations and the cpu time. Besides, to get a clear observation, we give the horizontal coordinate, a log-scale.

The numerical results show the efficiency comparisons, with some other three-term conjugate gradient methods.

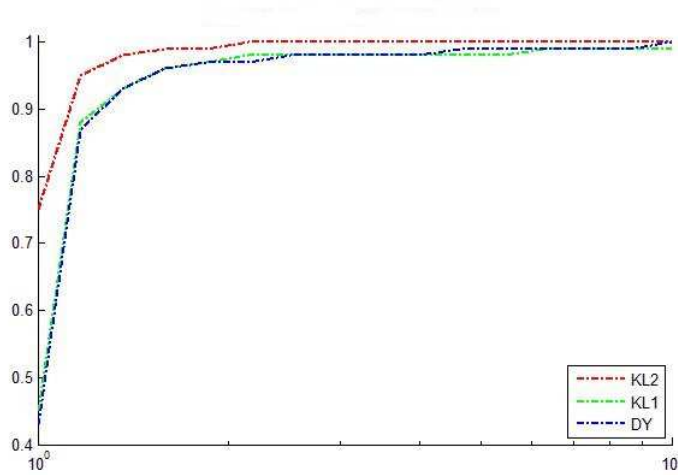


Figure 1: Comparison Based on Iteration

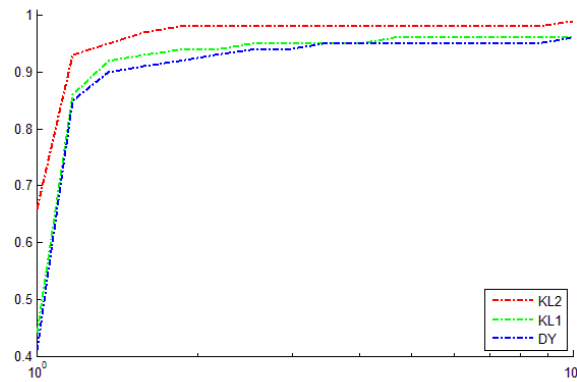


Figure 2: Comparison Based on Function

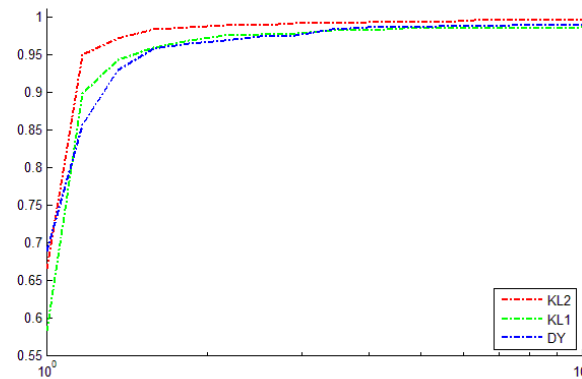


Figure 3: Comparison Based Ontime

REFERENCES

1. Abbas Y. Al-Bayati, Abdul-Ghafoor J. and Khalil A. (2009) "Two- Version of conjugate gradient algorithm Based on conjugacy Conditions for unconstrained optimization". *American J. of Economics and Business Administration* 1, (2).
2. Abbo
3. Andrei N. (2008), "An Unconstrained Optimization test function collection". *Adv. Model. Optimization*. 10. pp.147-161.
4. Dai Y.H. and Yuan Y., "A nonlinear conjugate gradient method with a strong global convergence property", *SIAM Journal on Optimization*, 10 (1999), 177– 182.
5. Dai,Y. and Liao, L. (2001) "New conjugacy conditions and related non-linear CG methods" *Appl. Math optim.*,(43). Springer-verlag, New York.
6. Dolan. E. D and J. J. Mor'e, "Benchmarking optimization software with performance profiles", *Math. Programming*, 91 (2001), pp. 201-213.
7. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.
8. Fletcher R., "Practical Methods of Optimization", 1: Unconstrained Optimization, JohnWiley& Sons, New York, (1987).
9. Hameed, M. Sadiq., (2013)," Modified Conjugate Gradient Algorithms For Solving Unconstraint Minimization Problems" M.Sc. Thesis, College of Computer Sciences And Mathematics, University of Mosul.
10. Hestenes M. R. and Stiefel E., "Methods of conjugate gradient for solving linear systems," *Journal of Research of the National Bureau of Standards, Sec. B* 48:409–436, (1952).
11. Perry, A. (1978), "A modified conjugate gradient algorithms", *Operations Research*, (26), 1073–1078.
12. Polak E., and Ribiere, G., (1969), Not sur la convergence de directions conjugue'e. *Rev. Franaise Informant Rechercheoperationelle, 3e Anne'e.*, Vol. 16, pp. 35-43.
13. Tomizuka. H and Yabe. H, (2004). "A Hybrid Conjugate Gradient method for unconstrained Optimization".
14. Wolfe P., "Convergence conditions for ascent methods", *SIAM Rev.*, 11: 226235, (1969).
15. Wolfe P., "Convergence conditions for ascent methods". II: Some corrections, *SIAM Rev.*,13: 185-188, (1971).
16. Zhang L., Zhou W. and Li, D.H. "Some descent three-term conjugate gradient methods and their global convergence",

Optimization Methods and Software, 22: 697–711(2007).

17. Zhang L., Zhou W. and Li, D.H. "Global convergence of modified Fletcher-Reeves conjugate gradient method with Armijo-type line search", *Numerical Math.*, 104(2006).